

DOCUMENT RESUME

ED 244 526

HE 017 227

AUTHOR Norman, G. R.
TITLE Standardized Patients.
PUB DATE Apr 84
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 23-27, 1984).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Clinical Teaching (Health Professions); Higher Education; *Medical Education; *Patients; Reliability; *Simulation; Validity
IDENTIFIERS *Simulated Patients

ABSTRACT

The use of healthy individuals acting as simulated patients for the purpose of clinical teaching is discussed. The term "standardized patients" is used to refer to training the individual to present a standard, repeatable stimulus. Some evidence suggests that simulated patients possess high fidelity (i.e., closely approximate real patients). The reliability of scores based on simulated patients depends on the scoring method. Inter-rater reliability based on a single patient encounter ranges from acceptable to very high, depending on the degree of objectivity present in the scoring system. The question of validity, as opposed to fidelity, is a correlational one: i.e., do individuals who perform well or poorly on simulated patient encounters also perform well or poorly on other measures of competence? Evidence appears to indicate that scores derived from simulated patient encounters primarily reflect interpersonal skills, and that these scores are unrelated to content knowledge. A consistent finding has been that students taught with simulated patients out-performed those taught with real patients. A major barrier to the extensive use of standardized patients is feasibility (cost of training and application, use of faculty resources, and acceptability by faculty). (SW)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

☒ This document has been reproduced as
received from the person or organization
originating it.
☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

D. K.
Norman

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

STANDARDIZED PATIENTS

G.R. NORMAN

Although healthy individuals acting as patients have always had a role in the teaching of the clinical examination, Barrows (1971) was the first to formally train individuals to simulate all aspects of a patient problem—history, physical findings, and affect. Initially simulated patients were trained to simulate neurological problems; since that time a wide variety of problems in most specialties of medicine have been accurately simulated. Although many clinicians are skeptical about the fidelity of such simulations, usually only a brief exposure is necessary to convince even the most cynical of the realism of live simulated patients.

Simulated patients are purported to have a number of advantages over actual patients; they can be scheduled in advance and chosen for their appropriateness to the learning objectives, they can simulate acute, serious or rare conditions which might not usually be encountered in wards or clinics, free discussion can be conducted in the presence of the patient, under circumstances which would be viewed as unethical or impolite with real patients, junior students can approach and examine simulated patients without concern for physical or emotional harm, and simulated patients can provide accurate and objective feedback without fear of consequences.

Stillman (1980) has extended this approach in a format she calls

ED244526

NE 017 227

"patient instructors". She uses real patients with chronic conditions who are extensively trained to simulate their manner at the time of the first encounter with a health professional, and then to provide detailed feedback to the examiner regarding the conduct of the examination. In order to encompass both these approaches, we will use the term "standardized patients" implying a training of the individual to present a standard, repeatable stimulus.

-FIDELITY-

As we have mentioned, much anecdotal evidence exists to suggest that simulated patients possess high fidelity--i.e. closely approximate real patients. Several studies have verified this claim.

Burri, et al (1976) introduced trained simulated patients into family physician's offices in the Hamilton area; he reported that none were detected by the study physicians. Owen and Winkler (1974) also used standardized patients to assess quality of care delivered by family physicians in Australia; of the patients none were detected.

A study conducted by our research team (Norman and Tugwell, (1982)) focussed more directly on behavior with simulated and real patients. Ten residents each saw four real patients and four simulated patients presenting the same problem. Essentially no differences were observed in history-taking, physical examination, and diagnosis. Residents confronted with both real and simulated patients correctly identified 66 per cent against a chance accuracy of 50 per cent, and many

indicated that they would not have been able to succeed if they had only seen a single presentation.

A more recent unpublished study at McMaster using methods similar to those of Burri and Reynard introduced patients into the offices of practising physicians in a town in southern Ontario. Twenty per cent of the patients were identified, primarily because of one bizarre problem. It is not known, however, how many real patients were thought by these physicians to be simulated!

-RELIABILITY-

It is not terribly relevant to describe the reliability of simulated patients directly, since, to use a behaviourist analogy, the simulated patient represents the stimulus and it is the reliability of the responses that is most problematic. In passing, it would appear that simulated patients do give a fairly reliable or reproduceable stimulus, although this would be a difficult question to address rigorously because patient encounters are not scripted and some variability would be expected.

What of the reliability of scores based on simulated patients? Not surprisingly, it depends on the scoring method. Subjective ratings by observers using categories such as "Completeness of History" and five point scales, yield inter-observer reliability coefficients typically in the range 0.5 to 0.7, (Finkel and Norman, (1972)). Many other approaches have been used to score simulated patient encounters,

however, often with very high reliability. As one example, the studies of Elstein et al (1978) and Barrows et al (1982) used simulated patients to investigate clinical reasoning, and coded twenty to thirty variables from each encounter with high reliability (but dubious validity as measures of competence!). We have also developed a method to score the outcome of a simulated patient encounter, focussing on the differential diagnosis and list of investigations Norman, (1984). The score uses weights for each alternative based on performance of a criterion group, and constructs a ratio of the weights of options selected by the subject to the weights of an ideal list of options. The method is almost "objective", therefore highly reliable. Another approach has been the application of quality of care approaches to simulated patient encounters. Because the patient problem is standardized, a criterion group can determine (at least in theory) those options on history, physical, etc which would constitute an appropriate workup of the problem. We have demonstrated agreements better than 95 per cent between recall of the simulated patient and observer rating of these checklists. ^(Kappan = .86) ⁽¹⁹⁸⁰⁾ Stillman reported similarly — eg Stillman 1980. high reliability, with Phi coefficients ranging from .77 to .89.

No doubt other scoring methods with other scales have been, or will be, used. Nevertheless, a reasonable conclusion is that inter-rater reliability based on a single patient encounter ranges from acceptable to very high, depending on the degree of objectivity present in the scoring scheme.

That is the good news. The dilemma is that, in most applications, the intent is not to obtain a reliable rating of performance on that day,

with that patient, dressed in those clothes, etc. Rather, it is hoped that the patient encounter represents a "biopsy" of the physician's competence. As such, one would hope that scores based on simulated patient encounters would be correlated across encounters.

Simply put, they are not. Killer (1984) has shown a correlation across seven GI problems of 0.16. A study we reported last year (Norman (1983)) showed correlations ranging from around 0.3 for two presentations of the same problem to 0.1 for problems in different specialties. Similar results were reported by Rutala (1980). *Edelman*
a range of correlations from -.26 to +.22.

This dilemma, which has been labelled (or mislabelled) "content specificity", appears to be a problem common to nearly all evaluations based on a single patient encounters, and further discussion will follow.

-VALIDITY-

The question of validity, as opposed to fidelity, is a correlational one; i.e. do individuals who perform well or poorly on simulated patient encounters also perform well or poorly on other measures of competence? Although the two issues are inter-related, they are not "isomorphic". For instance although, on average, subjects may perform exactly the same with simulated patients as with real patients, which appears to be the conclusion from the review of the fidelity section, it is possible that no correlation exists between performance in the two settings. Conversely, a method may have low fidelity, yet

individuals who perform well or poorly in the test situation may continue to perform well or poorly in real life.

The issue of "content-specificity" also represents a real constraint to any demonstration of validity, since a stable estimate of performance based on simulated patients will require a large number of encounters.

One study, (Wakefield and Norman (1977)), compared performance of family medicine residents based on an average of twenty-two observed encounters to their performance assessed concurrently by supervisor reports and later performance on the certification examination. Scores on data gathering were uncorrelated with either criterion. Clinical judgement was correlated .51 with the same competency assess by supervisors, and .54 with "Treatment" assessed at certification by a formal oral. Interpersonal and interview skills were correlated about .3 to .4 with similar categories assess by supervisor report and three oral examinations on the certification examination.

Rutala (1980) found positive correlations between scores of "content" assessed by patient instructors and peer ratings in data acquisition, problem recognition, attitudes, and competence, again supporting the concurrent validity of simulated patients.

Two studies, (Killer, et al (1984) and Norman, et al (1983)) have examined the correlation between simulated patient performance and knowledge assessed by multiple choice tests; neither demonstrated any significant correlation.

Thus it would appear that scores derived from simulated patient encounters primarily reflect interpersonal skills, and the evidence suggests that these scores are unrelated to content knowledge.

-LEARNING-

In contrast to the paucity of studies of reliability and validity, there are numerous studies showing positive educational effects of simulated patients in such domains as inter-personal skills, physical examination, and pelvic examination. In particular, several studies have contrasted learning from simulated and real patients, (Tinning (1975); Livingstone (1978); Holzman (1977); Johnson (1976), Helfer (1975); and Maquire (1977)). A consistent finding from all studies is that students taught with simulated patients out performed those taught with real patients, on a variety of measures from satisfaction to neurological examination skills.

-FEASIBILITY-

A major barrier to more extensive use of standardized patients is undoubtedly one of feasibility, which reduces to several issues cost of training and application, use of faculty resources, and acceptability by faculty.

It would seem that the cost of training and application is consistently overestimated. At McMaster, simulated patients are trained in 2 to 3

hours, at a cost of \$10/hour, a total cost which would be the envy of any computer simulation or PMP designer. Cost of application is again about \$10/hour, which compares favourably with the cost of, for example, computer simulations. The real cost issue is one of faculty time—observation of simulated patients inventors requires extensive use of faculty resources. However, there has been significant progress towards the use of this resource for teaching and evaluation without the requirement for faculty input, which may contribute greatly to its dissemination. Perhaps as faculty are exposed to simulated patients, and to the studies which indicate a strong positive effect on learning, and acceptable reliability and validity, this simulation method will be more extensively used.

REFERENCES

- 1 Barrows, H. S. Simulated Patients. Springfield, Ill. Charles C. Thomas, 1971.
- 2 Stillman, P. L., et al. Patient Instructors as Teachers and Evaluators. J. Med. Educ. 55, 186-193, 1980.
- 3 Burri, A, et al. Using Simulated Patients to Evaluate Practising Physicians in a Community. Proc. 15th Conf. Res. Med. Educ., San Francisco, Ca, 1976.
- 4 Owen, A., Winkler, R. General Practitioners and Psycosocial Problems: An Evaluation Using Pseudopatients. Med. J. Australia 2, 393-398, 1974.
- 5 Norman, G. R., Tugwell, P. A Comparison of Resident Performance on Real and Simulated Patients. J. Med. Educ. 57, 708-715, 1982.
- 6 Finkel, A. and Norman, G. R. The Validity of Direct observation. Proc 12th Conf Res. Med. Educ. Washington, D.C., 1973.
- 7 Elstein, A. S., Shulman, L. S., Sprofska, S. A. Medical Problem Solving: An Analysis of Clinical Reasoning. Cambridge, Massachusetts: Harvard University Press, 1978.

- 8 Barrows, H. S., et al. The Clinical Reasoning of Randomly Selected Physicians in General Medical Practice. Cln. Invest. Med. 5, 49-55, 1982.
- 9 Norman, G. R. Objective Measurement of Clinical Performance, Med. Educ. (under review) 1984.
- 10 Killer, D., Tugwell, P., Norman, G.R.
- 11 Norman, G. R., et al. The Generalizability of Measures of Clinical Problem-Solving. Proc. 22nd Conf. Res. Med. Educ. Washington, D.C., 1983.
- 12 Rutala, P.J., Stillman, P.L., Sabers, D.L. Patient Instructors as Evaluators of Housestaff's Clinical Competence. Proc. 19th Conf. Res. Med. Educ., Washington, D.C., 1980.
13. *Stillman PL, Ruggill JS, Rutala PJ, Sabers DL Patient instructors as teachers & evaluators J Med Educ 55, 186-191, 1980.*
- 13 Wakefield, J.G., Norman, G.R. Assesment of Cognitive and Interpersonal Skills in Clinical Problem Solving. Proc. 16th Conf. Res. Med. Educ., Washington, D.C., 1977.
- 14 Tinning, F.C. Simulation in Medical Education. Ph.D. Thesis. Michigan State University, 1975.

- 15 Livingstone, R., Ostrow, D.N. Professional Patient Instructors in the Teaching of the Pelvic Examination. A. J. Obstet. Gynecol. 132, 54-57, 1978.
- 16 Holzman, G. B., et al. Initial Pelvic Examination Instruction: The Effectiveness of Three Contemporary Approaches. Am. J. Obstet. Gynecol. 129, 124-129, 1977.
- 17 Johnson, C. F., et al. Sick Infants Versus Simulated Well-Baby Examination as an Initial Pediatric Learning Experience. J. Med. Educ. 51, 1021-1023, 1976.
- 18 Helfer, R. E., et al. A Comparison of Pediatric Interviewing Skills Using Real and Simulated Mothers. Pediatrics 55, 397-400, 1975.
- 19 Maquire, G.P., Clerlce, D., Jolley, B. An Experimental Comparison of Three Courses in History-Taking Skills for Medical Students. Med. Educ. 11, 175-182, 1977.